DOCUMENT RESUME

ED 328 072 FL 019 037

AUTHOR Read, John
TITLE The Validity of Writing Test Tasks.
PUB DATE Apr 90
NOTE 16p.; Paper presented at the Regional Language Centre Seminar (Singapore, April 9-12, 1990).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS English for Academic Purposes; Foreign Countries; Higher Education; Language Proficiency; *Language Tests; *Prior Learning; Second Language Instruction; Second Languages; *Test Format; *Test Items; *Test Validity; Writing Exercises; *Writing Tests
IDENTIFIERS *Victoria University of Wellington (New Zealand)

ABSTRACT

This paper, a discussion of the use of written tests to assess second language proficiency and achievement, considers what constitutes a valid writing test task and addresses three questions: (1) To what extent is performance influenced by prior knowledge about the topic? (2) Does it make a difference how the writing task is specified on the test paper? and (3) Do different types of tasks produce significant differences in the performance of learners in a writing test? Three task types varying in the amount and nature of content material provided as part of task specification are distinguished: independent tasks, guided tasks, and experience tasks. A test developed at the English Language Institute of Victoria University (New Zealand) and administered to foreign students before they begin study consists of three tasks, one guided, one guided/experience, and one clearly an experience task. It is proposed that while classification by task type focuses attention on one important dimension of writing test tasks--namely, the extent to which they provide content support for test-takers--other factors in task validity must be considered, including the amount of guidance given concerning the quantity and form of text expected, the clarity of task definition, the choice of general versus field-specific topic, and assessment focus on product rather than the process of writing. Several approaches to increasing writing test validity are suggested. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED328072

THE VALIDITY OF WRITING TEST TASKS

John Read
English Language Institute
Victoria University of Wellington
New Zealand

Plenary paper presented at the RELC Regional Seminar on Language Testing and Language Programme Evaluation

SEAMEO Regional Language Centre
Singapore
9-12 April 1990

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. Read

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

FL019037

THE VALIDITY OF WRITING TEST TASKS

John Read
Victoria University of Wellington
New Zealand

There is a long tradition in the academic world of using essays and other forms of written expression as a means of assessing student proficiency and achievement. In earlier times essay writing seemed to be quite a straightforward method of examining student performance. However, with the development of the modern science of language testing, writing has come to be considered one of the more difficult skills to test adequately because we now recognise the importance of achieving a satisfactory level of both intra-rater and inter-rater reliability in the marking of such tests. It is no longer considered acceptable to rely simply on the subjective judgement of a single teacher who has not been specifically trained or guided for the task -- although it has to be admitted that this is an idea that dies hard in the context of academic assessment in the university.

The modern concern about achieving reliability in marking has meant that relatively less attention has been paid to the other major issue in the testing of writing: how to elicit samples of writing from the students. Recent research on writing involving both native speakers and second language learners raises a number of questions about the setting of writing test tasks. The relevant research involves not only the analysis of writing tests but also more basic studies of the nature of the writing process. Some of the questions that arise are as follows:

1 To what extent is performance influenced by the amount of prior knowledge that writers have about the topic that they are asked to write about in a test?

2 Does it make a difference how the writing task is specified on the test paper?

3 Do different types of task produce significant differences in the performance of learners in a writing test?

The purpose of this paper is to explore these questions, with particular reference to the author's experience in testing English for academic purposes, and then to consider the more general issue of what constitutes a valid writing test task.

The Role of Prior Knowledge

One starting point in the selection of writing test tasks is a consideration of the role that knowledge of the subject matter might play in writing performance. In the case of reading, it is now widely accepted - on the basis of the research by Carrell (e.g. 1984, 1987), Johnson (1981) and others - that background knowledge is a very significant factor in the ability of second language readers to comprehend a written text. Furthermore, testing researchers such as Alderson and Urquhart (1985) and Hale (1988) have produced some evidence that a lack of relevant background knowledge can affect performance on a test of reading comprehension. Until recently, there have been few comparable studies of the role of background knowledge in second language writing, but it seems reasonable to expect that it does have a similar effect: someone is likely to write better about a familiar topic than an unfamiliar one.

Of course, in general terms this factor has long been recognised as a significant one in the testing of writing, and there are various ways in which testers have sought to minimise its effect. One approach is to give careful consideration to the choice of topic. Jacobs et al. (1981:12-15) suggest, among other things, that the topic should be appropriate to the educational level and interests of the students, it should motivate them to communicate with the reader and should not be biased in favour of any particular sub-group among them. In other words, it should be a subject about which all potential test-takers have enough relevant information or opinions to be able to write to the best of their ability. On the other hand, it should not be too simple or predictable.

Another solution is to give the test-takers a choice of topics. In this case, it is assumed that there is a range of interests and backgrounds represented among the test-takers, and so it is hoped that all of them will find at least one that motivates them to write as best they can.

However, an alternative approach to the problem of the effect of background knowledge is to design tasks that provide the test-takers with relevant content material to work with. Thus, although differences in prior knowledge of the topic are not eliminated, the students are all provided with subject matter to use in completing the writing task, so that the focus of their efforts is not so much on generating ideas but more on expressing the ones provided in an appropriate manner.

A Classification of Writing Test Tasks

In order to provide a basis for analysing writing test tasks according to the amount of content material that they provide, it is useful to refer to Nation's (1990)

classification of language learning tasks. In this system, tasks are categorised according to the amount of preparation or guidance that the learners are given. If we adapt the classification to apply to the testing of writing, there are three task types that are relevant and they may be defined for our purposes as follows:

1 Independent tasks: The learners are set a topic and expected to write on it without any guidance.

This approach to the assessment of writing, known as the timed impromptu test, is commonly used in universities in the United States, especially in large-scale placement and proficiency tests. It assumes that all of the test-takers have background knowledge in the form of information, ideas and opinions that are relevant to the topic set. Another term which describes this type of task is "free composition".

In their simplest form, independent writing tasks can be illustrated by means of these topics, which are typical of those which are used in the composition section of the Michigan English Language Assessment Battery (MELAB):

The role of agriculture in my country today

Why young people in my country need a college education

Meeting the energy needs of a modern world - problems and prospects

(Quoted in Jacobs, Zingraf et al., 1981)

2 Guided tasks: The learners are provided with guidance while they are writing, in the form of a table, a graph, a picture or relevant language material.

Here we are not referring to "guided composition", in which lower proficiency learners are given language support, but rather tasks which provide content support, especially in the form of non-linear text material.

One major test which uses guided tasks of this second kind is the Test of Written English, the direct writing component of TOEFL (Test of English as a Foreign Language). In one of the two alternating versions of this test, the candidates are presented with data in the form of a graph or a chart and are asked to write an interpretation of it (Educational Testing Service, 1989: 9). For example, a preliminary version of the test included three graphs showing changes in farming in the United States from 1940 to 1980, the task being to explain how the graphs were related and to draw conclusions from them.

3 Experience tasks: The students are given the opportunity to acquire relevant content and skills through prior experience before they undertake the writing task.

Tasks of this kind are found in major EAP tests such as the International English Language Testing Service (IELTS) and the Test in English for Educational Purposes (TEEP). In both of these tests, writing tasks are linked with tasks involving other skills to some extent in order to simulate the process of academic study. For example, in the first paper of TEEP, the candidates work with two types of input on a single topic: a lengthy written academic text and a ten-minute lecture. In addition to answering comprehension questions about each of these sources, they are required to write summaries of the information presented in each one (Associated Examining Board, 1984). The same kind of test design, where a writing task requires the synthesizing of information from readings and a lecture presented previously on the same topic, is found in the Ontario Test of ESL (OTESL) in Canada (Wesche, 1987).

Thus, the three types of task vary according to the amount and nature of the content material that is provided to the test-takers as part of the task specification. The assumption is that this may help to reduce the effects of differences in background knowledge among test-takers and, when the writing tasks are linked to earlier reading and listening tasks, may represent a better simulation of the process of academic study than simply giving a stand-alone writing test.

The Tasks in the ELI Writing Test

In order to illustrate in practical terms the use of guided and experience tasks in the assessment of writing for academic purposes, let us look at the tasks used in a writing test developed at the English Language Institute of Victoria University. The test is administered at the end of a three-month EAP course for foreign students preparing for study at New Zealand universities, and forms part of a larger proficiency test battery. The test results provide a basis for reporting on the students' proficiency to sponsoring agencies (where applicable) and to the students themselves. They are not normally used for university admission decisions; other measures and criteria are employed for that purpose.

The test is composed of three tasks, as follows:

Task 1

The first task, which is modelled on one by Jordan (1980: 49),

is an example of the guided type. The test-takers are given a table of information about three grammar books. For each book the table presents the title, author name(s), price, number of pages, the level of learner for whom the book is intended (basic, intermediate or advanced) and some other features, such as the availability of an accompanying workbook and the basis on which the content of the book is organised. The task is presented to the learners like this: "You go to the university bookshop and find that there are three grammar books available. Explain which one is likely to be the most suitable one for you by comparing it with the other two."

This is a guided task in the sense that the students are provided with key facts about the three grammar books to refer to as they write. Thus, the focus of their writing activity is on selecting the relevant information to use and organising the composition to support the conclusion that they have drawn about the most suitable book for them.

Task 2

For the second task, the test-takers are given a written text of about 600 words, which describes the process of steel-making. Together with the text, they receive a worksheet, which gives them some minimal guidance on how to take notes on the text. After a period of 25 minutes for taking notes, the texts are collected and lined writing paper is distributed. Then the students have 30 minutes to write their own account of the process, making use of the notes that they have made on the worksheet but not being able to refer to the original text.

It could be argued that this second task is another example of the guided type, in the sense that the students are provided with a reference text that provides them with content to use in their writing. However, it can also be seen as a simple kind of experience task, because the test procedure is divided into two distinct stages: first, reading and notetaking and then writing. While they composing their text, the students can refer only indirectly to the source text through the notes that they have taken on it.

Task 3

The third task, like the second one, is intended to simulate part of the process of academic study. In this case, the preparation for the test begins five days beforehand, because in the week leading up to the test the students all study the topic on which the test task is based as part of their regular classwork. The topic used so far has been Food Additives. The classes spend about five hours during the week engaged in such activities as reading relevant articles, listening to mini-lectures by the teacher, taking notes, having a class debate and discussing how to organise an answer to a specific

question related to the topic. However, the students do not practise the actual writng of the test task in class.

The week's activities are intended to represent a kind of mini-course on the topic of Food Additives, leading up to the test on the Friday, when the students are given an examination-type question related to the topic, to be answered within a time limit of 40 minutes. The question is not disclosed in advance either to the students or the teachers. A recent question was as follows:

> Processed foods contain additives.
> How safe is it to eat such foods?

This third task in the test is a clear example of an experience task. It provides the students with multiple opportunities during the week to learn about the topic (to acquire relevant prior knowledge, in fact), both through the class work and any individual studying they may do. Of course this does not eliminate differences in background knowledge among the students on the course. However, it is considered sufficient if the students' interest in the topic is stimulated and they learn enough to be able to write knowledgeably about it in the test.

Towards a Broader Analysis of Test Tasks

The classification into independent, guided and experience types focuses attention on one important dimension of writing test tasks: the extent to which they provide content support for the test-takers. However, recent developments in the study and teaching of writing have highlighted a variety of other considerations that need to be taken into account, and it is these that we now turn.

The literature on academic writing for native speakers emphasizes the need to state explicitly some of the requirements that have traditionally been taken for granted. It is now more widely recognised that students need not only the ability to marshal content material effectively but also to tailor their writing for specific readers (or "audiences") and purposes. In addition, they need guidance on such matters as the amount to be written and form of text that they should produce. If such specifications are needed for native speakers of English, then the need is even greater in the case of foreign students who - as Hamp-Lyons (1988: 35) points out - often lack knowledge of the discourse and pragmatic rules that help to achieve success in an academic essay test.

At a theoretical level, Ruth and Murphy (1984) developed a model of the "writing assessment episode" involving three actors - the test-makers, the test-takers and the test-raters - and correspondingly three stages: the preparation of

the task, the student's response to it and the evaluation of the student's response. The model highlights the potential for mismatch between the definition of the task as intended by the test-makers and as perceived by the students. Assuming that in a large-scale testing programme the test-raters are different people from the test-makers, the way that the raters interpret the task is a further source of variability in the whole process.

The tasks in the ELI writing test described earlier can be used to illustrate some of the problems of interpretation that arise. For example, in Task 1, the purpose of the task is not very clear. In real life, if one were making a decision about which grammar book to buy, one would normally weigh up the various considerations in one's mind or at most compose a list similar to the table in the task specification, rather than writing a prose text about it. Some of the students who took the test apparently had this difficulty and their solution was to compose their response in the form of a letter, either to a friend or two the manager of the university bookstore. Neither of these alternatives was entirely satisfactory, but the fact that some students responded in this way represents useful feedback on the adequacy of the task specification. The instinct of some of the teacher-raters was to penalize the letter-writers, on the grounds that there was nothing in the specification of the task to suggest that a letter should be written. However, one can equally argue that it was the task statement that was at fault.

Another kind of interpretation problem has arisen with the third task, when the question has been stated as follows:

> Processed foods contain additives.
> How safe is it to eat such foods?

The test-setter intended the first sentence to be an introductory statement that could be taken as given; it was the second sentence that was supposed to be the actual writing stimulus. However, in the most recent administration of the test, a number of the students took it as a proposition to be discussed and devoted the first half of their composition to it, before moving on to the question of the safety of the additives. Ruth and Murphy (1984: 417-418) noted the same phenomenon with a similar writing prompt for L1 students in the United States. Whereas the majority of the students considered the opening sentence of the prompt to be an operative part of the specification that needed to be referred to in their essays, none of the teacher-raters thought it necessary to do so.

Faced with such problems of interpretation, test-writers are in something of a dilemma. In a writing test, it is obviously desirable to minimize the amount of time that the students need to spend on reading the question or the task statement. On the other hand, as Hamp-Lyons (1988: 38)

points out, a short statement may give inadequate guidance to the students on the kind of composition that they are expected to produce. This suggests that longer statements are necessary, but also that they must be composed with great care in order to reduce the possibility of misinterpretation.

The following example of a "state-of-the-art" writing task specification comes from a proposed test prepared by the U.S. Defense Language Institute for the selection of language teaching applicants:

> . Assume that you have just returned from a trip and are writing a letter to a close friend. Describe a particularly memorable experience that occurred while you were traveling.
> . This will be one paragraph in a longer letter to your friend. The paragraph should be about 100 words in length.
> . You will be judged on the style and organization of this paragraph as well as vocabulary and grammar. Remember, the intended reader is a close friend. (Herzog, 1988: 155)

In terms of our classification, this is still an independent task, because it provides no content material for the test-takers to work with, but obviously it provides explicit specifications for the task in other respects.

Carlson and Bridgeman (1986: 139-142) give a useful summary of the factors to be considered in designing a stimulus for an academic writing test, based on their experience with the development of the Test of Written English. However, the most comprehensive system for classifying writing tasks is undoubtedly that established by Purves and his colleagues (Purves, Soter, Takala and Vahapassi, 1984) for the IEA Study of Written Composition in secondary schools in eighteen countries. Their system consists of fifteen dimensions: instruction, stimulus, cognitive demand, purpose, role, audience, content, rhetorical specification, tone/style, advance preparation, length, format, time, draft, and criteria. Obviously not all of these dimensions need to be specified in any individual writing stimulus, but they highlight the complexities involved, especially when the test-setters and the test-takers do not share the same cultural and educational backgrounds.

Does Type of Task Make a Difference?

A more general issue related to the preparation of writing tests is whether the type of task makes a difference to the students' performance. No matter how carefully a task is specified, its validity may be limited if it does not provide a basis for making generalizations about the test-

takers' writing ability. The issue is of particular relevance for a test like the Test of Written English (TWE) because, although two different task types are used, only one of them is included in any particular administration of the test. Thus, if task type is a significant variable, candidates may be advantaged or disadvantaged depending on which version of the test they take.
Carlson and Bridgeman (1986) report that the pilot study of the TWE showed no significant differences in performance on the two types of task. However, Stansfield and Ross (1988) acknowledge that this is a question which requires further investigation.

In fact, Reid (1988) found evidence of differences between the TWE tasks when she analysed student scripts from a pre-operational version of the test using the Writer's Workbench, a computer text-analysis program which provided data on discourse fluency, lexical choice and the use of cohesion devices in the students' writing. The program revealed significant differences in the discourse features of the texts produced in response to the two different tasks.

There are a number of other recent research studies which provide evidence that type of task does make a difference in tests for both L1 and L2 writers. In their study of L1 secondary students in Scotland, Pollitt and Hutchinson (1987) used five different tasks. They found that the easier tasks were those in which the content and organization of the text were cued in various ways by the test stimulus. In addition, tasks that were closer to personal experience and spoken language (letter-writing or story-telling) were less demanding than more formal ones, like expressing an opinion on a controversial topic.

This variation according to formality was also present in Cumming's (1989) research on L2 writers at a Canadian university. The ratings of his three tasks differed significantly and, in particular, there was a clear distinction between the informal topic (a personal letter) and the more academic ones (an argument and a summary).

Of course, in a test of academic writing, one would noramlly expect that only more formal tasks would be set. However, a recent study by Tedick (1988) indicates that there is another distinction that is relevant to the setting of academic writing tests: that of general vs. specific topics. Tedick's subjects, who were graduate ESL students in the United States, wrote one essay on a general topic and another on a topic related to their own fields of study. The essays on the field-specific topic were longer, more syntactically complex and of higher overall quality (as measured by holistic ratings) than the essays on the general topic. Furthermore, the field-specific essays provided better discrimination of the three levels of ESL proficiency represented among the subjects. Tedick concluded that allowing students to make use of prior knowledge of their academic subject gave a better

measure of their writing proficiency than the kind of general topic that is commonly used in writing tests.

These findings can be interpreted in relation to Douglas and Selinker's concept (1985) of "discourse domains", which are the content areas that are particularly meaningful and important for individual test-takers. Like Tedick, Douglas and Selinker argue that performance in tests of language production is affected by whether the test-takers can relate the test topic to their own interests and fields of specialization. Although these authors have looked specifically at oral production, it seems reasonable to expect that discourse domains may play a role in writing test performance as well.

There is clearly more research to be done to explore the variability in types of writing task. As Stansfield and Ross (1988) point out in a survey of research needs for the Test of Written English, there are two ways in which writing tasks can be shown to be meaningfully different. The first kind of evidence is psychometric: are there significant differences in the rankings of the test-takers when the ratings from each of the tasks are subjected to correlational analysis? In other words, are the different tasks measuring the same underlying construct? Secondly, one can use linguistic evidence to identify differences by means of syntactic and discourse analyses of the students' texts. Thus, measures of fluency, frequency of error and syntactic complexity can be obtained to reflect various aspects of the quality of the writing. Stansfield and Ross argue that, from the point of view of construct validity, the psychometric evidence is crucial: "if an empirical analysis of performance ratings on each task failed to show any significant variation between the two sets of ratings, one could claim that both were tapping the same construct, even if qualitative differences were found in the language used on each task." (op. cit.: 166) However, they acknowledge that at least some writing specialists consider that qualitative, linguistic differences are also important.

For the present, it seems prudent to include a variety of writing tasks in a writing test. It can be argued that this not only makes the assessment more reliable by producing several samples of writing from each student, but it contributes to the validity of the test by giving a broader basis for making generalizations about the student's writing ability.

Conclusion: Writing as Process

However, this leads to one final issue that is relevant to a consideration of the validity of writing tasks. The testing of writing inevitably focuses on the text that the test-taker produces or, in other words, the product rather

than the process. Practical constraints normally mean that the students can be given only a limited amount of time for writing and therefore they must write quite fast in order to be able to produce an adequate composition. The preceding discussion of different tasks and ways of specifying them has concentrated on the issue of how to elicit the kind of texts that the test-setter wants, with little consideration of the process by which the texts will be produced and whether that reflects the way that people write in real life.

However, any contemporary discussion of writing assessment must take account of the major developments that have occurred over the last fifteen years in our understanding of writing processes. In the case of L1 writers, researchers such as Britton et al. (1975), Murray (1978), Perl (1980) and Graves (1983) have demonstrated how people compose a segment of text, pause to read and consider it, revise or replace it, plan further segments of text and so on. Murray (1978) described writing as "a process of discovery", through which writers explored their ideas and found out what they wanted to express. As Kelly (1989: 80) puts it, "the act of writing has great generative power, both in the sense of creating ideas and creating the language to express those ideas". Studies by Zamel (1983), Raimes (1987), Cumming (1989) and others have found that L2 writers exhibit very much the same strategies in composing text as L1 writers do.

There are a number of implications of this research for the assessment of writing. If the production of written text is not strictly linear but inherently recursive in nature, normally involving cyclical processes of drafting and revising, this suggests that test-takers need to be given time for thinking about what they are writing; in addition, explicit provision needs to be made for them to revise and rewrite what they have written.

The question, then, is how the writing process can be accommodated within the constraints of the test situation. Of course, any test situation is different from the context of real-world language processing, but the disjunction between "natural" writing processes and the typical writing test is quite marked, particularly if we are interested in the ability of students to write essays, research reports and theses, rather than simply to perform in examination settings. Even a substantially increased time allocation for completing a test task does not alter the fact that the students are being required to write under constraints that do not normally apply to the writing process.

There are various ways in which one can reduce the effects of the time constraints. One means is to limit the demands of the writing test task by providing support as part of the task specification. The provision of content material, as is done with guided and experience tasks, is one way of reducing the complexity of the writing task and allowing the test-taker to focus on the structure and

Another, more radical approach which is gaining ground is to move away from a reliance on timed tests for writing assessment. This may not be possible in large-scale placement or proficiency tests for practical reasons, but there is now increasing interest in portfolio assessment (see, e.g., Katz, 1988: 196-198), which involves collecting a standard set of different types of writing completed by each student over a period of time (usually as part of a course) and then having them assessed according to agreed criteria by one or two teachers other than the class teacher. Once again, there are practical difficulties in implementing this approach, but it may be valuable as a supplementary method of assessment, especially in the case of postgraduate students whose primary academic writing activities are the preparation of theses and research papers.

Clearly, all of these considerations indicate that the validity of writing test tasks is a complex issue and one that is likely to receive increasing attention in the years to come.

REFERENCES

Alderson, J. C. and Urquhart, A. H. (1985). The effect of students' academic discipline on their performance on ESP reading tests. Language Testing, 2, 192-204.

Associated Examining Board. (1984). The Test in English for Educational Purposes. Aldershot, England: Author.

Britton, J. et al. (1975). The development of writing abilities 11-18. London: Macmillan.

Carrell, P. L. (1984). The effects of rhetorical organization on ESL readers. TESOL Quarterly, 18, 441-469.

Carrell, P. L. (1987). Content and formal schemata in ESL reading. TESOL Quarterly, 21, 461-481.

Carlson, S. and Bridgeman, B. (1986). Testing ESL student writers. In K.L. Greenberg, H.S. Wiener and R.A. Donovan (Eds.) Writing assessment: issues and strategies. New York: Longman.

Cumming, A. (1989). Writing expertise and second-language

proficiency. Language Learning, 39.1, 81-141.

Douglas D. and Selinker, L. (1985) Principles for language tests within the "discourse domains" theory of interlanguage: research, test construction and interpretation. Language Testing, 2.2, 205-226.

Educational Testing Service. (1989). Test of Written English guide. Princeton, NJ: Author.

Graves, D. (1983). Writing: children and teachers at work. London: Heinemann Educational Books.

Hale, G. A. (1988). Student major field and text content: interactive effects on reading comprehension in the Test of English as a Foreign Language. Language Testing, 5, 49-61.

Hamp-Lyons, L. (1988). The product before: Task-related influences on the writer. In P.C. Robinson (ed.) Academic writing: Process and product. (ELT Documents 129). London: Modern English Publications/British Council.

Herzog, M. (1988). Issues in writing proficiency assessment. Section 1: The government scale. In P. Lowe, Jr. and C. W. Stansfield (Eds.) Second language proficiency assessment: Current issues. Englewood Cliffs, NJ: Prentice Hall Regents.

Jacobs, H. L., Zingraf, S. A., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). Testing ESL composition: A practical approach. Rowley, MA: Newbury House.

Johnson, P. (1981). Effects on reading comprehension of language complexity and cultural background of a text. TESOL Quarterly, 15, 169-181.

Jordan, R. R. (1980). Academic writing course. London: Collins.

Katz, A. (1988). Issues in writing proficiency assessment. Section 2: The academic context. In P. Lowe, Jr. and C. W. Stansfield (Eds.) Second language proficiency assessment: Current issues. Englewood Cliffs, NJ: Prentice Hall Regents.

Kelly, P. (1989). Theory, research and pedagogy in ESL writing. In C.N. Candlin and T. F. McNamara (Eds.) Language, learning and community. Sydney: National Centre for English Language Teaching and Research, Macquarie University.

Murray, D. (1978). Internal revision: a process of discovery. In C. R. Cooper and L. Odell (Eds.) Research

on composing: Points of departure. Urbana, Ill.: NCTE.

Nation, P. (1990). A system of tasks for language learning. In S. Anivan (Ed.) Language teaching methodology for the Nineties. Singapore: SEAMEO Regional Language Centre.

Perl, S. (1980). Understanding composing. College Composition and Communication, 31.4, 363-369.

Pollitt, A. and Hutchinson, C. (1987) Calibrating graded assessments: Rasch partial credit analysis of performance in writing. Language Testing, 4.1, 72-92.

Purves, A. C., Soter, A., Takala, S., and Vahapassi, A. (1984). Towards a domain-referenced system for classifying composition assignments. Research in the Teaching of English, 18.4, 385-416.

Raimes, A. (1987). Language proficiency, writing ability and composing strategies: A study of ESL college writers. Language Learning, 37.3, 439-467.

Reid, J. M. (1988) Quantitative differences in English prose written by Arabic, Chinese, Spanish and English students. Unpublished Ph.D. dissertation, Colorado State University.

Ruth, L. and Murphy, S. (1984). Designing topics for writing assessment: Problems of meaning. College Composition and Communication, 35.4, 410-422.

Smith, W. L., Hull, G. A. and others. (1985). Some effects of varying the structure of a topic on college students' writing. Written Communication, 2, 73-89.

Stansfield, C. W. and Ross, J. (1988). A long-term research agenda for the Test of Written English. Language Testing, 5, 160-186.

Tedick, D. J. (1988). The effects of topic familiarity on the writing performance of non-native writers of English at the graduate level. Unpublished Ph.D. dissertation, Ohio State University.

Wesche, M. B. (1987). Second language performance testing: The Ontario Test of ESL as an example. Language Testing, 4, 28-47.

Zamel, V. (1983). The composing processes of advanced ESL students: Six case studies. TESOL Quarterly, 17, 165-187.